

# An analytic framework for stochastic binary biological switches

Guilherme C.P. Innocentini

*Departamento de Matemática Aplicada,*

*Instituto de Matemática e Estatística,*

*Universidade de São Paulo, São Paulo, Brazil*

Sarah Guiziou, Jerome Bonnet

*CBS, CNRS UMR 5048 - UM - INSERM U1054, Montpellier, France.*

Ovidiu Radulescu

*DIMNP, UMR CNRS 5235 - University of Montpellier, Montpellier, France.*

(Dated: June 15, 2016)

We propose an analytic solution for the stochastic dynamics of a binary biological switch, defined as a DNA unit with two mutually exclusive configurations, each one triggering the expression of a different gene. Such a device could be used as a memory unit for biological computing systems designed to operate in noisy environments. We discuss a recent implementation of an exclusive switch in living cells, the recombinase addressable data (RAD) module. In order to understand the behavior of a RAD module we compute the exact time dependent distributions of the two expressed genes starting in one state and evolving to another asymptotic state. We consider two operating regimes of the RAD module: fast and slow stochastic switching. The fast regime is “aggregative” and produces unimodal distributions, whereas the slow regime is “separative” and produces bimodal distributions. Both regimes can serve to prepare pure memory states when all cells are expressing the same gene. The slow regime can also separate mixed states by producing two sub-populations each one expressing a different gene. Our model provides a simplified, general phenomenological framework for studying biological memory devices and our analytic solution can be further used to clarify theoretical concepts in bio-computation and for optimal design in synthetic biology.

PACS numbers:

Biological machines are a promising new paradigm in computation [11]. By using synthetic biology it is now possible to design memory units and logic gates operating in living cells [6, 7]. Contrarily to their silicon counterparts, biological computing units (BCU) can evolve their computation capacities by proliferation and auto-organization. Furthermore, parallel computing with large number of BCU could be used to performing distributed computation in living environments for medical applications such as implants for augmenting capacities or for health monitoring [4].

However, there are prices to pay when replacing the silicon substrate with biological cells. BCUs are submitted to stochastic fluctuations that are ubiquitous in the realm of molecular machines [8, 10]. Within cell communities, the result of a computation can vary from one cell to another and therefore this result should be represented as a probability distribution between the different states of the system. In order to optimize the design of BCUs, precise calculations of time dependent population distributions is extremely useful. In this paper we show how to solve this problem in the context of the set and reset of a reversible and stochastic, rewritable biological unit.

Synthetic biological memory devices have been engineered based on different mechanisms. Some systems use feedback loop in order to achieve data storage via bistability. For example, double negative feedback systems inspired by the phage lambda Cro/CI circuit and using two repressor mutually repressing each others have been implemented in the bacterium *Escherichia coli* [9, 12]. Positive feedback based memory devices have also been engineered in the baker's yeast *Sacharomyces cerevisiae* [2]. Recently, genetic memory systems using recombinases from the integrase family, enzymes that catalyze strand exchange between specific DNA sequences [4], have been implemented into living cells. DNA data storage has the advantage that only two discrete states are possible. The state of the system can be interrogated by direct DNA sequencing but also by measuring the intensity of a fluorescent protein (FP) whose expression is controlled by the invertible DNA sequence (called the DNA register). Fluorescent reporters of two colors, for instance Green Fluorescent Protein (GFP) and Red Fluorescent Protein (RFP), can be used to signify binary DNA register states of single cells. Although the majority of DNA data storage devices are single-write units, a particular architecture called the recombinase addressable data (RAD) module using two proteins, an integrase and a recombination directionality factor (or excisionase) enables

rewritable digital data storage in bacteria [5] (Fig.1).

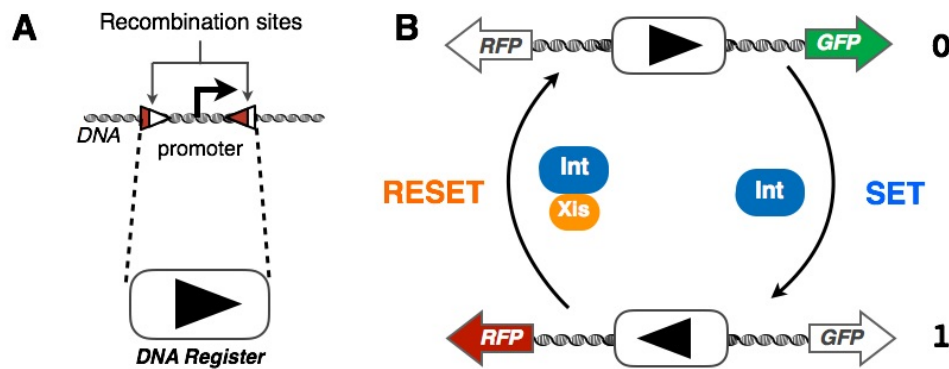


FIG. 1: Schematic architecture and operational principle of the Recombinase Addressable Data (RAD) module. A. The DNA register is composed of a DNA sequence containing a promoter (driving transcription) flanked by the two recombination sites recognized by the integrase (Int). B. In state 0, the promoter drives expression of GFP. Integrase catalyzes the SET reaction, inverting the DNA register and enables transition towards state 1, in which the promoter has been inverted and now drives transcription of RFP. Concomitant expression of Integrase and Excisionase (Xis) catalyze the RESET reaction by enabling transition from state 1 to state 0.

The RAD module proposed by Bonnet et al [5] is composed of a DNA sequence which can exist in two configurations inverted one with respect to the other. The transition between the configurations is reversible. Depending on the configuration, single cells produce one of the two reporters, a GFP or RFP (Fig.1). Although the two internal states of the memory defined by the DNA configurations are mutually exclusive, the expression of the two reporter genes depends on how fast the unit switches between the two configurations. As we will prove later in this letter, fast switching between the two configurations allow that both GFP and RFP are expressed by the same cell. In this work, we are interested in the conditions enabling the generation of cell populations residing in a pure state (when all the cells express the same reporter gene) or separated mixed state (when sub-populations express an unique reporter gene but not necessarily the same). To this end, we introduce a mathematical model to describe the stochastic dynamics of the system. Our model is based on two coupled master equations governing the evolution of the joint probability distributions carrying the information about switch status and numbers of GFP ( $g$ ) and RFP ( $r$ ) at a given instant of

time:

$$\begin{aligned} \frac{d\phi_{g,r}^0(t)}{dt} &= k_g[\phi_{g-1,r}^0(t) - \phi_{g,r}^0(t)] + \rho[(g+1)\phi_{g+1,r}^0(t) - g\phi_{g,r}^0(t)], \\ &+ \rho[(r+1)\phi_{g,r+1}^0(t) - r\phi_{g,r}^0(t)] - h\phi_{g,r}^0(t) + f\phi_{g,r}^1(t) \end{aligned} \quad (1)$$

$$\begin{aligned} \frac{d\phi_{g,r}^1(t)}{dt} &= k_r[\phi_{g,r-1}^1(t) - \phi_{g,r}^1(t)] + \rho[(r+1)\phi_{g,r+1}^1(t) - r\phi_{g,r}^1(t)]. \\ &+ \rho[(g+1)\phi_{g+1,r}^1(t) - g\phi_{g,r}^1(t)] + h\phi_{g,r}^0(t) - f\phi_{g,r}^1(t) \end{aligned} \quad (2)$$

The random variables are  $g$  and  $r$ . Production of GFP and RFP is controlled by the rates  $k_g$  and  $k_r$ , respectively. Degradation/dilution rate of both reporters is given by  $\rho$  and the switching between the two states is encoded in the rates  $h$  (SET) and  $f$  (RESET).

Introducing the generating functions,  $\phi^0(z, y, t) = \sum_{g,r=0}^{\infty} \phi_{g,r}^0(t) z^g y^r$  and  $\phi^1(z, y, t) = \sum_{g,r=0}^{\infty} \phi_{g,r}^1(t) z^g y^r$ , we transform the master equations in a set of partial differential equations:

$$\begin{aligned} \frac{\partial \phi^0}{\partial t} &= (z-1)[k_g \phi^0 - \rho \frac{\partial \phi^0}{\partial z}] - (y-1)\rho \frac{\partial \phi^0}{\partial y} - h\phi^0 + f\phi^1, \\ \frac{\partial \phi^1}{\partial t} &= (y-1)[k_r \phi^1 - \rho \frac{\partial \phi^1}{\partial y}] - (z-1)\rho \frac{\partial \phi^1}{\partial z} + h\phi^0 - f\phi^1. \end{aligned} \quad (3)$$

The desired joint probability distributions are obtained from the generating functions by applying the formulas:

$$\phi_{g,r}^0(t) = \frac{1}{r!} \frac{1}{g!} \frac{\partial^r}{\partial y^r} \frac{\partial^g}{\partial z^g} \phi^0(z, y, t) \Big|_{z=y=0}, \quad \phi_{g,r}^1(t) = \frac{1}{r!} \frac{1}{g!} \frac{\partial^r}{\partial y^r} \frac{\partial^g}{\partial z^g} \phi^1(z, y, t) \Big|_{z=y=0}. \quad (4)$$

In order to solve the system of PDEs in Eq.(3) we propose a set of transformations, that will lead to an integrable system of ODEs. To do so, we perform a first change of variables:

$$w = \frac{(z-1)}{(y-1)}, \quad x = (z-1)(y-1). \quad (5)$$

In the new set of variables  $(w, x)$  the equations read:

$$\begin{aligned} \frac{\partial \phi^0}{\partial t} + 2x\rho \frac{\partial \phi^0}{\partial x} &= \sqrt{wx} k_g \phi^0 - h\phi^0 + f\phi^1 \\ \frac{\partial \phi^1}{\partial t} + 2x\rho \frac{\partial \phi^1}{\partial x} &= \sqrt{\frac{x}{w}} k_r \phi^1 + h\phi^0 - f\phi^1. \end{aligned} \quad (6)$$

Note that this transformation let time ( $t$ ) unchangeable and had the effect of eliminating one of the partial derivatives. To eliminate another partial derivative and obtain a set of

ordinary differential equation with respect to the remaining variable we propose a second transformation,

$$\nu = \sqrt{\frac{x}{w}} \quad \mu = \sqrt{\frac{x}{w}} e^{-\rho t}, \quad (7)$$

which leads to the set of ordinary differential equations in the variable  $\nu$ :

$$\begin{aligned} \nu \rho \frac{\partial \phi^0}{\partial \nu} &= \nu w k_g \phi^0 - h \phi^0 + f \phi^1 \\ \nu \rho \frac{\partial \phi^1}{\partial \nu} &= \nu k_r \phi^1 + h \phi^0 - f \phi^1. \end{aligned} \quad (8)$$

Now, solving the first equation of the system (8) for  $\phi^1(\nu, \mu, w)$ , gives

$$\phi^1(\nu, \mu, w) = \frac{1}{f} \left( \nu \rho \frac{\partial \phi^0}{\partial \nu} - \nu w k_g \phi^0 + h \phi^0 \right), \quad (9)$$

and by substituting the result in the second equation of the system, we arrive to the second order ordinary differential equation with respect to the variable  $\nu$  for  $\phi^0(\nu, \mu, w)$  :

$$\nu \rho^2 \frac{\partial^2 \phi^0}{\partial \nu^2} + \rho [f + h + \rho - \nu(k_r + k_g w)] \frac{\partial \phi^0}{\partial \nu} - [w k_g (\rho + f) + k_r h - \nu w k_g k_r] \phi^0 = 0 \quad (10)$$

This equations has a regular singularity at  $\nu = 0$  and a irregular one at infinity. This structure suggest solutions in terms of confluent hypergeometric functions. To make it more clear, let us use the ansatz  $\phi^0(\nu, \mu, w) = \exp(\nu w k_g / \rho) \psi(\nu, \mu, w)$ , and a last transformation of variables:  $\nu = \eta \rho / (k_r - w k_g)$ . Putting all together, we arrive to the equation in the new variable  $\eta$  for  $\psi(\eta, \mu, w)$ :

$$\eta \frac{\partial^2 \psi}{\partial \eta^2} + (b - \eta) \frac{\partial \psi}{\partial \eta} - a \psi = 0, \quad (11)$$

where:  $a = h/\rho$ ,  $b = (h + f + \rho)/\rho$  and  $\eta = \nu(k_r - w k_g)/\rho$ .

Now, Eq.(11) is in the canonical form of the confluent hypergeometric equation, or Kummer equation, and the general solution is straightforward:

$$\psi(\eta, \mu, w) = F(\mu, w) M(a, b, \eta) + G(\mu, w) \eta^{1-b} M(1 + a - b, 2 - b, \eta). \quad (12)$$

where M stands for Kummer function,  $F(\mu, w)$  and  $G(\mu, w)$  are arbitrary functions that will be determined by the initial conditions.

To obtain the expressions for the generating functions we have to multiply the solution in Eq.(12) by the exponential factor,  $\exp(\nu w k_g / \rho)$ , to obtain  $\phi^0(\eta, \mu, w)$  using this result and Eq.(9) we can obtain the generating function for  $\phi^1(\eta, \mu, w)$ , in the variables  $(\eta, \mu, w)$ :

$$\begin{aligned} \phi^0(\eta, \mu, w) &= e^{\left(\frac{k_g w \eta}{k_r - k_g w}\right)} [F M(a, b, \eta) + G \eta^{1-b} M(1 + a - b, 2 - b, \eta)], \\ \phi^1(\eta, \mu, w) &= e^{\left(\frac{k_g w \eta}{k_r - k_g w}\right)} \left[ F \frac{h}{f} M(1 + a, b, \eta) - G \eta^{1-b} M(2 + a - b, 2 - b, \eta) \right]. \end{aligned} \quad (13)$$

The last task remaining is to find  $F(\mu, w)$  and  $G(\mu, w)$ . As said before, these functions are determined by the initial conditions. To accomplish that, one can see that setting  $t = 0$  in the second transformation Eq.(7) leads to  $\nu = \mu$ , which in the variables  $(\eta, \mu, w)$  implies in  $\eta = \tilde{\eta} = \mu(k_r - wk_g)/\rho$ . Specifying the functions  $\phi^0(\eta = \tilde{\eta}, \mu, w)$  and  $\phi^1(\eta = \tilde{\eta}, \mu, w)$  that will appear in the left hand side of Eq.(13) one can find the expressions for  $F(\mu, w)$  and  $G(\mu, w)$ . To do so, let us use vector and matrix notation to express the solutions in Eq.(13) as:  $\vec{\phi} = U\vec{F}$ , where,  $\vec{\phi} = (\phi^0(\eta, \mu, w), \phi^1(\eta, \mu, w))^T$ ,  $\vec{F} = (F(\mu, w), G(\mu, w))^T$  and the entries of the matrix  $U$  are given by:

$$\begin{aligned} U_{1,1} &= \exp\left(\frac{k_g w \eta}{k_r - k_g w}\right) M(a, b, \eta), \\ U_{1,2} &= \exp\left(\frac{k_g w \eta}{k_r - k_g w}\right) \eta^{1-b} M(1 + a - b, 2 - b, \eta), \\ U_{2,1} &= \exp\left(\frac{k_g w \eta}{k_r - k_g w}\right) \frac{h}{f} M(1 + a, b, \eta), \\ U_{2,2} &= -\exp\left(\frac{k_g w \eta}{k_r - k_g w}\right) \eta^{1-b} M(2 + a - b, 2 - b, \eta). \end{aligned} \tag{14}$$

Inverting  $\vec{\phi} = U\vec{F}$  we obtain the expression  $\vec{F} = U^{-1}\vec{\phi}$ . Setting  $\eta = \tilde{\eta}$  bring us to the position of determining the vector  $\vec{F} = (F, G)^T$  through the initial conditions  $C_0 = \phi^0(\eta = \tilde{\eta}, \mu, w)$  and  $C_1 = \phi^1(\eta = \tilde{\eta}, \mu, w)$ . One final observation, before presenting  $F$  and  $G$ , concerns the determinant of the matrix  $U$  which is necessary because we need this determinant ( $\det(U) = U_{1,1}U_{2,2} - U_{1,2}U_{2,1}$ ) to compute  $U^{-1}$  (the inverse matrix of  $U$ ). The inspection of (14) reveals that  $\det(U)$  is a product of Kummer functions with a exponential envelope. At a first glance, no problem arise, however, as we will deal with the inverse of matrix  $U$  the inverse of this determinant is required. Due to the well known properties of the Kummer functions [1] this determinant assumes a very simple formula:

$$\det(U) = \frac{(b-1)e^{\frac{\eta(k_r - wk_g)}{k_r - k_g w}} \eta^{1-b}}{1 + a - b},$$

which is used to compute  $U^{-1}$ . The explicitly expressions for  $F$  and  $G$  are:

$$\begin{aligned} F &= \exp\left(\frac{-k_r \tilde{\eta}}{k_r - wk_g}\right) \frac{f}{h+f} [M(2 + a - b, 2 - b, \tilde{\eta})C_0 + M(1 + a - b, 2 - b, \tilde{\eta})C_1], \\ G &= \exp\left(\frac{-k_r \tilde{\eta}}{k_r - wk_g}\right) \tilde{\eta}^{b-1} \left[ \frac{h}{h+f} M(a+1, b, \tilde{\eta})C_0 - \frac{f}{h+f} M(a, b, \tilde{\eta})C_1 \right]. \end{aligned} \tag{15}$$

Where initial conditions are encoded in  $C_0 = \phi^0(\eta = \tilde{\eta}, \mu, w)$  and  $C_1 = \phi^1(\eta = \tilde{\eta}, \mu, w)$ .

At this point we are in position to exhibit the biological features of our model encoded in the time dependent joint distributions of three variables: the two-valued DNA register state and the numbers of GFP ( $g$ ) and RFP ( $r$ ). However, before doing that, let us rephrase our parameter space in adimensional biological terms. The two numbers  $N_g = k_g/\rho$  and  $N_r = k_r/\rho$  are the protein production efficiencies. The asymptotic occupancy probabilities are  $p_0 = f/(f+h)$  and  $p_1 = h/(f+h)$ . We call *switching flexibility* the important parameter  $\epsilon = (h+f)/\rho$ , representing the sum of the frequencies of the SET and RESET transitions. The two situations  $\epsilon < 1$  and  $\epsilon > 1$  correspond to slow and fast switching, respectively.

Although we have the time dependent solutions at hand we will first illustrate the role of the biological parameters by analyzing their influence in the shape of the asymptotic distributions of the model. The generating function for the the steady-state distributions is obtained by performing the limit  $t \rightarrow \infty$  in Eqs.(13), resulting in:

$$\begin{aligned}\phi^0(\eta, w) &= p_0 \exp\left(-\frac{N_g w \eta}{N_g w - N_r}\right) M(a, b, \eta), \\ \phi^1(\eta, w) &= p_1 \exp\left(-\frac{N_g w \eta}{N_g w - N_r}\right) M(1+a, b, \eta),\end{aligned}\tag{16}$$

where, in the biological parameter space we have:  $a = p_1 \epsilon$ ,  $b = \epsilon + 1$ , and  $\eta = \nu(N_r - w N_g)$ . Applying formulas (4) we obtain the steady-state joint probability distributions as

$$\begin{aligned}\phi_{g,r}^0 &= \frac{f}{f+h} \frac{\Delta^{g+r} e^{-\Delta}}{r!g!} \sum_{s=0}^g \binom{g}{s} (-1)^s \frac{(a)_s (a+s)_r}{(b)_s (b+s)_r}, \\ \phi_{g,r}^1 &= \frac{h}{f+h} \frac{\Delta^{g+r} e^{-\Delta}}{r!g!} \sum_{s=0}^g \binom{g}{s} (-1)^s \frac{(a+1)_s (a+1+s)_r}{(b)_s (b+s)_r},\end{aligned}\tag{17}$$

where we have used  $N_g = N_r = \Delta$  and the symbol  $(\bullet)_s$  is the Pochhammer's symbol [1].

With Eqs.(17), we have computed the total joint probability distribution in the asymptotic regime of the system for different values of  $p_0, p_1$  and  $\epsilon$  keeping  $\Delta$  constant. The steady-state distributions are represented in Figs.2a-c. In order to expand the regions of small expression we made a change of variables in the distributions from  $g, r$  variables to  $\log(g), \log(r)$  variables. As said before, we are doing this exercise to show the influence of the parameters in the shape of the distributions and, moreover, to distinguish between mixed and pure states of the DNA register. This concept is important for the relation between hidden memory states and visible (readable) phenotype. Let us define a mixed state by  $0 < p_0 < 1$  and  $0 < p_1 < 1$ , whereas a pure state means that either  $p_0 = 1$  or  $p_1 = 1$ . Also,

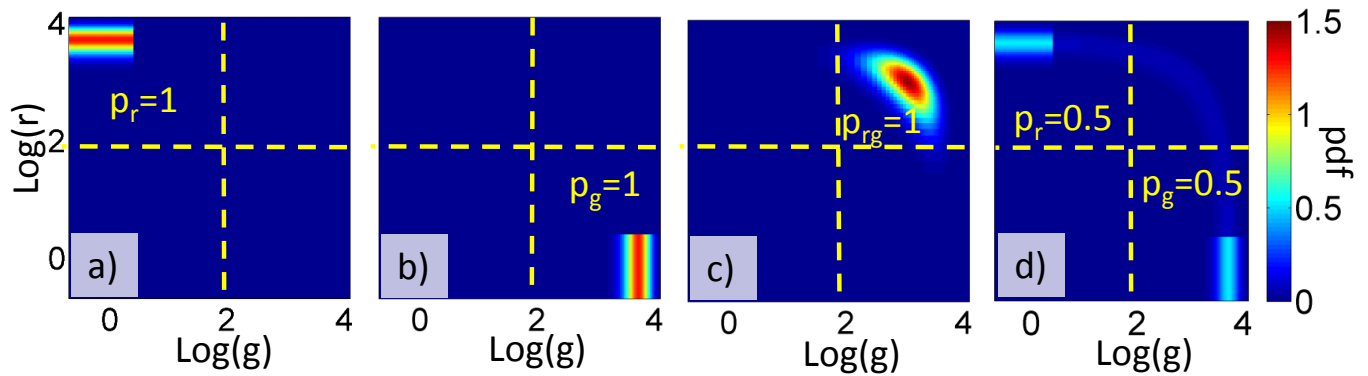


FIG. 2: Computed steady state distributions of GFP and RFP for a) pure  $p_1 = 1$  red state (all cells express RFP) and  $\epsilon = 0.5$  (slow switching); b) pure  $p_1 = 1$  green state (all cells express GFP)  $\epsilon = 0.5$  (slow switching); c) mixed  $p_0 = p_1 = 0.5$  unimodal state (all cells express both GFP and RFP) and  $\epsilon = 10$  (fast switching); d) mixed  $p_0 = p_1 = 0.5$  bimodal state (half of the cells express only GFP and half only RFP) and  $\epsilon = 0.5$  (slow switching). In logarithmic variables, the double integral of the probability distribution is normalized to one. The remaining parameters are:  $\Delta = 40$  and  $\rho = 1$ . The probabilities  $p_r$ ,  $p_g$  and  $p_{rg}$  indicate the fractions of cells with  $\log(g) < 2, \log(r) > 2$ ,  $\log(g) > 2, \log(r) < 2$  and  $\log(r) > 2, \log(g) > 2$ , respectively, where the threshold 2 corresponds to the middle of the expression dynamical interval.

we define three types of sub-populations regarding the level of expression of: GFP only; RFP only and both GFP and RFP. The corresponding state occupancies are  $p_g$ ,  $p_r$ ,  $p_{rg}$ , respectively. A well separated population has low values of  $p_{rg}$ . The state occupancies are not necessarily the same as the occupancy probabilities  $p_0$  and  $p_1$  encoding the probability to find the DNA register in state 0 or 1, respectively. A RAD unit in a pure state has a pure phenotype ( $p_g = 1$  or  $p_r = 1$ ), in other words all the bacteria express GFP in state 0 or RFP in state 1, as show in Fig.2a,b. The readout of a mixed state will depend on the switch flexibility. A fast switch ( $\epsilon > 1$ ) will correspond to unimodal distribution of expressed genes, Fig.2c (each bacteria expressing both genes in different proportions) whereas a slow switch ( $\epsilon < 1$ ) corresponds to bimodal population, Fig.2d, where some cells are green and others are red.

In order to illustrate the dynamical behavior of the RAD module we will first set the initial conditions ( $C_0$  and  $C_1$ ) of the system. To do so, we use the steady-state solutions as initial conditions but with different values for the occupancy probabilities and switch



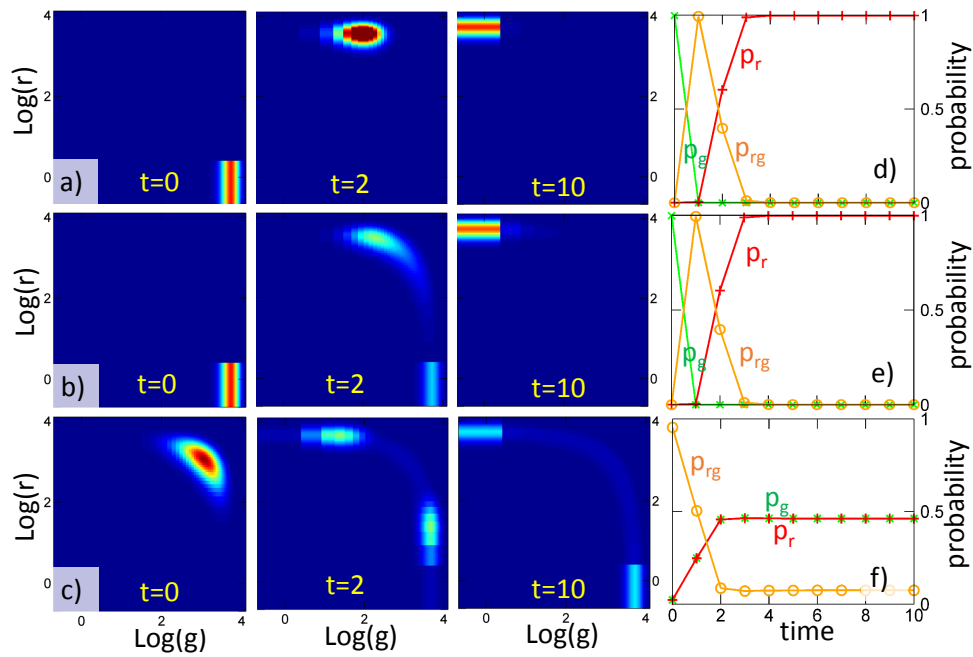


FIG. 3: Time dependent distributions of a RAD unit (time is measured in units of  $\rho^{-1}$ ). In a), b) we illustrate setting of a pure state. The  $t = 0$  state corresponds to a pure distribution with  $\tilde{p}_0 = 1$  (green) and switch flexibility  $\tilde{\epsilon} = 0.5$ . During set operation (for  $t > 0$ ) the unit evolves towards a pure state  $p_1 = 1$  (red) but with different switch parameters:  $\epsilon = 10$  for a) (fast switching) and  $\epsilon = 0.5$  for b) (slow switching). In c) we illustrate separation of a unimodal mixed state. An unimodal mixed state with  $\tilde{p}_0 = \tilde{p}_1 = 0.5$  and  $\tilde{\epsilon} = 10$  (fast switching) is transformed to a bimodal mixed state by lowering the switch flexibility,  $\epsilon = 0.1$  (slow switching). Also, in c) one can see that for fast switching the phenotype is unimodal: bacteria express both GFP and RFP. The  $t = 10$  phenotype, is bimodal: 50% of the cells express RFP and 50% express GFP. The remaining parameters are:  $N_g = N_r = 40$  and  $\rho = 1$ . The rightmost columns d-f) show the time dependence of the state occupancies.

flexibility  $(\tilde{p}_0, \tilde{p}_1, \tilde{\epsilon})$ , leading to the initial generating functions:

$$\begin{aligned}\phi^0(\tilde{\eta}, w) &= \tilde{p}_0 \exp\left(-\frac{N_g w \tilde{\eta}}{N_g w - N_r}\right) M(\tilde{a}, \tilde{b}, \tilde{\eta}), \\ \phi^1(\tilde{\eta}, w) &= \tilde{p}_1 \exp\left(-\frac{N_g w \tilde{\eta}}{N_g w - N_r}\right) M(1 + \tilde{a}, \tilde{b}, \tilde{\eta}),\end{aligned}\tag{18}$$

At  $t = 0$  we abruptly change the occupancy probabilities and switch flexibility from  $(\tilde{p}_0, \tilde{p}_1, \tilde{\epsilon})$  to  $(p_0, p_1, \epsilon)$ , keeping them constant for  $t \geq 0$ . To study set and reset dynamics of the RAD

unit we consider three distinct experiments and we show the corresponding time dependent joint probability distributions for each one of these experiments in Fig.3. The first two experiments correspond to setting a pure state. We start with initial condition corresponding to a pure state  $\tilde{p}_0 = 1$  where all cells are producing GFP (in the setting described in Fig.1 this means that a strong excisionase signal is applied together with the integrase), and at time  $t = 0$  we change the asymptotic occupancy parameter to the complementary pure state,  $p_1 = 1$ , driving all the cells to be producing RFP in the asymptotic configuration (excisionase is washed-out). During the setting we use fast switching ( $\epsilon = 10$ ) for the first experiment Fig.3(a) and slow switching ( $\epsilon = 0.5$ ) for the second one Fig.3(b) (corresponding to high and low concentrations of integrase). In the third experiment, Fig.3(c), we start with an unimodal mixed steady-state configuration, in the fast switch regime ( $\epsilon = 10$ , corresponding to high integrase and excisionase concentrations) and change the switch flexibility to low values ( $\epsilon = 0.1$ , lower concentrations). We call this last experiment “developing” because it transforms the initial unimodal mixed state in which single cells express both GFP and RFP (state occupancy  $p_{rg} = 1$ ) into a bimodal, “separated” mixed state when two sub-populations have pure phenotypes, expressing either RFP or GFP ( $p_r = p_g \approx 0.5$  in Fig.2). More generally, the final state occupancies are given by the DNA occupancy probabilities  $p_r \approx p_1, p_g \approx p_0$ . Therefore, lowering the switching frequencies by lowering the integrase and excisionase concentrations reveals previously hidden information about the DNA register occupancy probabilities. The last column of Fig.3 shows the time dependence of the state occupancies ( $p_g(t)$ ,  $p_r(t)$  and  $p_{rg}(t)$ ) corresponding to each one of the three experiments.

These experiments emphasize a clear distinction between slow and fast switches. Slow switches are “separative”, they tend to transform unimodal distributions into bimodal ones. They can be used as developers of mixed states, as occupancy probability can be read on bimodal distributions by counting cells in the two sub-populations, whereas it is much more difficult to estimate it from unimodal distributions. This can also be seen from the values of the state occupancies  $p_{rg}(t)$  that remain low (indicating separated population) during the dynamics. For slow switches there is a trade-off between good separation and bandwidth because the time needed to reach a final unimodal, pure state distributions is longer. Fast switches are “aggregative”, they tend to transform bimodal distributions into unimodal distributions.

The analytic time dependent solutions for the master equations describing the RAD

memory unit can be used to quantitative and qualitatively establish the basis for the design of such biological devices. All the necessary information about the system is encoded in the time dependent joint probability distributions which are obtained by applying formulas (4) in the analytic solutions of the model. The solutions are expressed in closed form in terms of the well known Kummer functions and packages in softwares of symbolic computation, such as Maple, are available for direct computation of their series expansions leading to the desired joint distributions. If, instead of evaluating the derivatives of the solutions in  $z = y = 0$  as in formulas (4), one evaluates them at  $z = y = 1$ , closed expressions for any desired moment of the joint distributions, such as mean values, variance and covariance are easily obtained. A special word should be said about “developing” of a mixed state using slow switching. Although this procedure has no direct interest for storing information, it could be used as a biological implementation of a p-switch. A p-switch [13] is a stochastic element that takes two values, one and zero with probabilities  $p$  and  $1 - p$ , respectively. In other words, it is a physical implementation of a Bernoulli random variable, which is the basic element in stochastic computing [3], an hybrid, digital/analog form of computation in which real numbers are represented as random finite sequences of zeros and ones, the represented real number  $p$  being the frequency of occurrence of ones. In our case, each random bit in a sequence would be a cell. Although redundant from an information theoretic point of view, this representation was shown to be particularly robust and protected against random faults [3].

Although the discussion is based on the RAD switch setting illustrated in Fig.1, our model provides a simplified phenomenological framework for studying other implementations of two states memories, including toggle-switches based on positive feedback. In the latter case, simplified switching models corresponding to stochastic transitions between the two attractors of the system can be justified by low noise approximations as discussed in [10].

*Acknowledgments.* Work supported by FAPESP, SP, Brazil (G.I., contract 2012/04723-4) and CNPq, Brazil (G.I., contract 202238/2014-8). O.R. thanks LABEX Epigenmed for support.

- 
- [1] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1964.

- [2] Caroline M Ajo-Franklin, David A Drubin, Julian A Eskin, Elaine PS Gee, Dirk Landgraf, Ira Phillips, and Pamela A Silver. Rational design of memory in eukaryotic cells. *Genes & development*, 21(18):2271–2276, 2007.
- [3] Armin Alaghi and John P Hayes. Survey of stochastic computing. *ACM Transactions on Embedded computing systems (TECS)*, 12(2s):92, 2013.
- [4] Yaakov Benenson. Biomolecular computing systems: principles, progress and potential. *Nature Reviews Genetics*, 13(7):455–468, 2012.
- [5] Jerome Bonnet, Pakpoom Subsoontorn, and Drew Endy. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proceedings of the National Academy of Sciences*, 109(23):8884–8889, 2012.
- [6] Jerome Bonnet, Peter Yin, Monica E Ortiz, Pakpoom Subsoontorn, and Drew Endy. Amplifying genetic logic gates. *Science*, 340(6132):599–603, 2013.
- [7] Jennifer AN Brophy and Christopher A Voigt. Principles of genetic circuit design. *Nature methods*, 11(5):508–520, 2014.
- [8] Carlos Bustamante, Wei Cheng, and Yara X Mejia. Revisiting the central dogma one molecule at a time. *Cell*, 144(4):480–497, 2011.
- [9] TS Gardner, CR Cantor, and JJ Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339–342, Jan 20 2000.
- [10] Thomas B Kepler and Timothy C Elston. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophysical journal*, 81(6):3116–3136, 2001.
- [11] Sergi Regot, Javier Macia, Núria Conde, Kentaro Furukawa, Jimmy Kjellén, Tom Peeters, Stefan Hohmann, Eulàlia de Nadal, Francesc Posas, and Ricard Solé. Distributed biological computation with multicellular engineered networks. *Nature*, 469(7329):207–211, 2011.
- [12] Z Toman, Christine Dambly-Chaudière, Liliane Tenenbaum, and Miroslaw Radman. A system for detection of genetic and epigenetic alterations in escherichia coli induced by dna-damaging agents. *Journal of molecular biology*, 186(1):97–105, 1985.
- [13] Daniel Wilhelm and Jehoshua Bruck. Stochastic switching circuit synthesis. In *2008 IEEE International Symposium on Information Theory*, pages 1388–1392. IEEE, 2008.